ZIBELINE INTERNATIONAL™ PUBLISHING

Matrix Science Mathematic (MSMK)

DOI: http://doi.org/10.26480/msmk.02.2025.56.60

ISSN: 2521-0831 (Print)
ISSN: 2521-084X(Online)
CODEN: MSMAD

CrossMark

RESEARCH ARTICLE

AN ENHANCED CONJUGATE GRADIENT METHOD FOR SOLVING UNCONSTRAINED OPTIMIZATION PROBLEMS

Husein Saleem Ahmed*, Salah Gazi Shareef

Faculty of Science, University of Zakho, Kurdistan Region, Iraq.
****Corresponding Author e-mail: hosensalimramrom@gmail.com***

ARTICLE DETAILS

Article History:

Received 23 April 2025
Revised 18 May 2025
Accepted 08 June 2025
Available online 08 July 2025

ABSTRACT

We aimed to develop a conjugate gradient method by reformulating parameters so that iterative optimization techniques perform more effectively. Instead of using a set of basic conjugate gradient formulas, the methodology introduces a new parameter that produces better convergence. The method is applied in Fortran to test how many iterations and how many evaluations of the function are needed, as listed in table 1. The behavior of convergence and the results used for comparison are created with Matplotlib on Python and Ggplot2 on R programming for the chart. We like to compare our method to the proven LS (Liu-Storey) method when checking how effective our proposed method is. We found that the technique provides better results with lower iteration numbers and better convergence speeds for many test cases, proving it can challenge conventional methods in optimizing problems without constraints.

KEYWORDS

Conjugate gradient, Unconstrained optimization, Descent condition, Global convergence, new vector.

1. INTRODUCTION

In this project, we solve an unconstrained minimization problem stated in the form,

$$min_{x\in R^n} f(x), where\ f: R \to R^n, \tag{1}$$

It is a continuously differentiable nonlinear function. And its gradient is defined as

$$\nabla f(x) = \left[\frac{\partial f}{\partial_1}, \frac{\partial f}{\partial x_2}, \dots, \frac{\partial}{\partial x_n}\right]^T = g_i, \tag{2}$$

The Conjugate Gradient (CG) method, iterative for solving (1), takes the following form:

$$x_{i+1} = x_i + ad_i, fori\ \geq 0, \tag{3}$$

Where $x_{i+1} - x_i = v_i$ and d_i *is* Search direction states recursively as:

$$d_i = \begin{cases} -g_i & if\ i - 0 \\ -g_{i+B_i d_{i-1}} & if\ i > 0' \end{cases} \tag{4}$$

Where B_i It is the CG coefficient, and there are different formulas for determining it. B_i Each is related to a distinct conjugate gradient approach. Some well-known strategies for determining Bi include:

i. The researchers proposed one of the first CG methods for optimization (Hestenes et al., 1952). Where the parameter is defined as

$$B_i^{HS} = \frac{g_{i+1}^T y_i}{d_i^T y_i}, \tag{5}$$

ii. Fletcher and Reeves (FR) proposed a CG method (Fletcher et at., 1964). This method is especially effective in nonlinear problems, where the parameter is defined as:

$$B_i^{FR} = \frac{g_{i+1}^T g_{i+1}}{g_i^T g_I}, \tag{6}$$

iii. Polack-Ribiere-Polack (PRP) proposed an improved CG method, the parameter is defined as (Polak et al., 1969; Polyak et al., 1969):

$$B_i^{PRP} = \frac{g_{i+1}^T y_i}{g_i^T g_i}, \tag{7}$$

iv. Liu and Storey(LS)(Liu et al., 1991).

$$B_i^{LS} = \frac{g_{i+1}^T y_i}{-d_i^T g_i}, \tag{8}$$

v. Dai and Yuan (DY): This method is proposed to enhances the convergence properties by adjusting the search direction (Dai et al., 1999).

$$B_i^{DY} = \frac{g_{i+1}^T g_{i+1}}{d_i^T y_i}, \tag{9}$$

vi. Conjugate Descent (CD) (Fletcher, R. 2000).

$$B_i^{CD} = \frac{g_{i+1}^T g_{i+1}}{d_i^T g_i}, \tag{10}$$

In addition to these standard algorithms, there are many other approaches for computing the CG coefficient; for more information, see (Shareef, 2022; Ibrahim et al., 2019; Khatab et al., 2024).

Researchers typically perform either exact or inexact line research. The CG approach, such as the strong Wolf conditions, is used to find αi.

The strong Wolfe rules are a pair of inequalities that aim to balance sufficient decrease of the objective function and guarantee that the search direction maintains a descent direction. These requirements are essential for establishing worldwide convergence of CG approaches.

Quick Response Code	Access this article online	
	Website: www.matrixsmathematic.com	**DOI:** 10.26480/msmk.02.2025.56.60

Cite The Article: Husein Saleem Ahmed, Salah Gazi Shareef (2025). An Enhanced Conjugate Gradient Method For Solving Unconstrained Optimization Problems. *Matrix Science Mathematic, 9(2): 56-60.*

- **Armijo Condition (Sufficient Decrease Condition)**

$$f(x_{i+1} + a_i x_i) - f(x_i) \leq \delta_{k1} \alpha_i g_i^T d_i, \quad (11)$$

- **Curvature Condition**

$$|g_{i+1}^T d_i| \leq \delta_{k2} g_i^T d_i, \quad (12)$$

Where 0< $\delta k2$<δK1<1 are constants according to Li and Weijun (Nocedal et al., 1999).

Section 2 proves the descent condition, sufficient condition, and Global Convergence. Using a new technique, we show that these three conditions are satisfied. Section 3 presents numerical results highlighting improvements in the number of iterations (NOI) and function evaluations (NOF). Additionally, a comparative analysis with the usual method is presented to demonstrate the improved performance.

2. Derivation Of Beta-Parameter Using A New Vector

To derive a new CG parameter, we first define a new vector as follows:

$$y_i^* = (1-\varepsilon) g_{i+1} - \varepsilon g_i, where\ 0 < \varepsilon < 1 \quad (13)$$

Next, multiply both sides of (13) by g_{i+1}^T, yielding

$$g_{i+1}^T y_i^* = g_{i\_1}^T [(1-\varepsilon) g_{i+1} - \varepsilon g_i] \quad (14)$$

We multiply both sides of (4) by y_i, to obtain

$$d_{i+1}^T y_i = g_{i+1 T} y_i + B_i^{new1} d_i^T y_i \quad (15)$$

using the modified conjugate gradient ($d_{i+1}^T y_i = -t g_{i+1}^T V_i$), we obtain

$$-t g_{i+1}^T V_i = -g_{i+1} T_{y_i} + B_i^{new1} d_i T_{y_i,} \quad (16)$$

Substituting (14) in (16), we derive a new CG as follows:

$$B_i^{new1} = \frac{g_{i+1}^T [(1-\varepsilon) g_{i+1} - \varepsilon g_i] - t g_{i+1}^T V_i}{d_i^T y_i}, \quad (17)$$

We can write a new search direction by

$$d_{i+1}^{new} = -g_{i+1} + \left[\frac{g_{i+1}^T [(1-\varepsilon) g_{i+1} - \varepsilon g_i] - t g_i^T v_i}{d_i^T y_i}\right] d_i, \quad (18)$$

3. Descent And Sufficient Descent Conditions For The New Algorithm

Theorem 1: In the convergence analysis of (CG) methods, we say that the descent condition holds if, for each search direction d_{i+1}. The following inequality is satisfied. $g_{i+1}^T d_{i+1}^{new} \leq 0$. In both cases, exact line search (ELS) and inexact line search (ILS). Then the search direction di+1 with modified parameter B_i^{new}, of the conjugate gradient method.

Proof: When $i = 0, d_0 = -g_0\ SO\ g_0 T_{d_0} = -||g_0||2_{<}0.$

When i=i+1, we have

$$d_{i+1} = g_{i+1} + B_i new1_{d_i},$$

Multiply both sides of (18) by $g^t{}_{i+1}$ $on\ the\ lift$, we obtain

$$g_{i+1}^T d_{i+1}^{new} = -g_{i+1}^T g_{i+1} + \left[\frac{g_{i+1}^T \left[(1-\varepsilon) g_{i+1} - \varepsilon g_i\right] - t g_{i+1}^t v_i}{d_i^T y_i}\right] g_{i+1}^T d_i,$$

If we choose αI using exact line search (ELS), which requires $g_{i+1}^T d_i = 0$. We have

$$g_{i=1}^T d_{i+1} = -\|g_{i+1}\|^2$$

Thus, $g_{i+1}^T d_{i+1} \leq 0$.

However, when choosing the step size a_i using inexact line search (ILS), $g_{i+1}^T d_i \neq 0$

$$g_{i+1}^T d_{i+1} = -\|g_{i+1}\|^2 + \left[(1-\varepsilon) \frac{g_{i+1}^T g_{i+1}}{d_i^T y_i} + \varepsilon \alpha \frac{g_{i+1}^T d_i}{d_i^T y_i} - t\alpha \frac{g_{i+1}^T d_i}{d_i^T y_i}\right] g_i^T d_i,$$

$$g_{i+1}^T d_{i+1} = -\|g_{i+1}\|^2 + (1-\varepsilon) \frac{g_{i+1}^T g_{i+1}}{d_i^T y_i} (g_i^T d_i) - a(t-\varepsilon) \frac{(g_{i+1}^T d_i)^2}{d_i^T y_i}$$

Since $g_{i+1}^T d_i \leq d_i^T y_i$, so

$$g_{i+1}^T d_{i+1} \leq -\varepsilon \|g_{i+1}\|^2 - a(t-\varepsilon) d_i^T y_i, \quad (19)$$

It is clearly $g_{i+1}^T d_{i+1} \leq 0$, because $\varepsilon, a(t-\varepsilon)$ and $d_i^T y_i$ are positive

Thus, $g_{i+1}^T d_{i+1} \leq 0$. ■

Theorem 2: Suppose that the step size (a_i) satisfies (11) and (12), and the search direction $d_{i+1} said$. To satisfy the sufficient descent condition, if there exists a constant c>0 such that. $g_{i+1}^T d_{i+1} \leq -c\|g_{i+1}\|^2$ holds when $0 \leq i$.

Proof: From equation (19), we obtain

$$g_{i+1}^T d_{i+1} \leq -\varepsilon \|g_{i+1}\|^2 - a(t-\varepsilon) d_i^T y_i \leq 0,$$

$$g_{i+1}^T d_{i+1} \leq -\|g_{i+1}\|^2 \left(\epsilon + a(t-\varepsilon) \frac{d_i^T y_i}{\|g_{i+1}\|^2}\right),$$

Assume, $c = \epsilon + a(t-\varepsilon) \frac{d_i^T y_i}{\|g_{i+2}\|^2}$, and c > 0

Finally, we obtain $g_{i+1}^T d_{i+1} \leq -c\|g_{i+1}\|^2$. ■

3.1 The Global Convergence Analysis of the New Algorithm

It was shown in the global convergence analysis of the new nonlinear CG algorithm that, whenever the gradient is Lipschitz continuous and the sets have a bound, the algorithm will converge to a stationary point starting from any position.

Assumption (Ei)

(E1): The level set $\Omega = x \in \frac{R^1}{f(x)} \leq f(x_0) + \epsilon$. The function f is bounded below.

(E2): In a neighborhood η of Ω, f is continuously differentiable and its gradient is Lipschitz continuous, i.e., there exists a constant L > 0 such that.

$$\|g(x_{i+1}) - g(x_i)\| \leq L\|x_{i+1} - x_i\| for\ all\ x \in n \quad (20)$$

We can write (20) by $y_i \leq L v_i$, (21)

Preposition: Under the assumption (Ei) $of\ f$, there exists a constant γ ≥ 0 such that.

$$\|g_{i+1}\| \leq \gamma, \quad (22)$$

Lema 1. If $\sum_{k \geq 1} \frac{1}{\|d_i\|^2} = \infty$ then $\lim_{k \to \infty} inf \|g_i\| = 0$

There exists a constant. $\boldsymbol{\vartheta}$**≥0** such that for all x, y∈$\boldsymbol{\varphi}$

$$\left(g(x) - g(y)\right)^T (x-y) \geq \vartheta \|x-y\|^2 \quad (23)$$

If f is a uniformly convex function, we can write Eq. (23) as:

$y_i^T v_i \geq \vartheta \|v_i\|^2$ Or $y_i^T d_i \geq \alpha\vartheta \|d_i\|^2$, (24)

Theorem 3: Suppose the assumption (Ei) holds and that is a uniformly convex function. The new algorithm of the equation (18), which satisfies the descent condition and is obtained by the strong Wolfe conditions (11) and (12), satisfies the global convergence

i.e. $\lim_{k \to \infty} inf \|g_i\| = 0$

Proof: We can write equation (17) as:

$$|B_i{}^{new}| = \left|(1-\varepsilon) \frac{g_{i+1}^T g_{i+1}}{d_i^T y_i} - \varepsilon \frac{g_{i+1}^T g_i}{d_i^T y_i} + - \frac{t g_{i+1}^T v_i}{d_i^T y_i}\right|,$$

$$|B_i{}^{new}| \leq \left|(1-\epsilon) \frac{g_{i+1}^T g_{i+1}}{d_i^T y_i}\right| + \left|(at-\varepsilon) \frac{g_{i+1}^T d_i}{d_i^T y_i}\right|,$$

From (21), (22), (23), and (24), we get

$$|B_i{}^{new}| \leq (1-\epsilon) \frac{a\gamma^2}{L\|vi\|^2} + (at-\varepsilon) \frac{a\gamma}{L\|v_i\|}$$

Let $\|v_i\| = \|x_{i+1} - x_i\|$, D=$max\|x_{i+1} - x_i\|, forallx\epsilon R$

We can write eq (18) *by*:

$$\|d_{i+1}^{new}\| \leq \|g_{i+1}\| + \left[(1-\epsilon) \frac{a\gamma^2}{L\|vi\|^2} + (at-\varepsilon) \frac{a\gamma}{L\|v_i\|}\right] \|d_i\|$$

$$\|d_{i+1}^{new}\| \leq \gamma + (1-\epsilon) \frac{a\gamma^2}{LD} + (at-\varepsilon) \frac{\gamma}{L} = \varphi$$

Now, by the above lemma, if $\sum_{i \geq 1} \frac{1}{\|d_{i+1}^{new}\|^2} = \sum_{i \geq 1} \frac{1}{\varphi^2} = \infty$, then

$\lim_{k \to \infty} inf \|g_i\| = 0$. ■

3.2 Algorithm of the new technique

- Data: Let x_0 as the starting value and $x_0 \epsilon^{R^n}$, set $d_0 = -g_0, i = 0$.
- Phase 1. If $\|g_i\| = 0$, then stop; otherwise, proceed to Phase 2.
- Phase 2. Determine the length of the step a_i By Wolfe conditions (11), (12).

- Phase 3. Set $x_{i+1} = x_i + v_i$.
- Phase 4. Compute g_{i+1}, if$\|g_{i+1}\| \leq 10^{-6}$ then stop. Else, go to Phase 5.
- Phase 5. Determine B_i^{new} *and* d_{i+1}^{new} by using
- $d_{i+1}^{new} = -g_{i+1} + \left[\frac{g_{i+1}^T[(1-\varepsilon)g_{i+1} - \varepsilon_{g_i}] - t g_{i+1}^T v_i}{d_i^T y_i}\right] d_i,$
- Phase 6. If $\|g_{i+1}\|^2 \leq \frac{|g_{i+1}^T g_i|}{0.2}$ Go to Phase 2, else set. $i = i + 1$ And go to Phase 3.

4. Numerical Result

Both the New Method and the LS-CG method have been applied to a set of over ten functions, where their problem dimensions vary from 5 to 5000. Two metrics are used to see how the convergence is reached: the number of iterations (NOI) and the number of function calls or evaluations (NOF). Table 1 clearly shows that the New Method always takes less time and uses fewer functions than LS-CG, no matter the complexity of the problems.

To keep the results the same, we opted for specific settings.

t=1 and ε = 0.0155

Table 1: A comparative performance analysis between the new algorithm and the standard LS-CG method.

Function Test	Dimension	New technique NOI-NOF	LS method NOI-NOF
MIELE x_0= (2,2,2,2,1)	5	17-56	30-106
	100	17-56	37-138
	1000	17-56	44-172
	3000	22-80	51-208
	5000	22-80	51-208
SUM x_0=1	5	6-39	6-39
	100	13-73	14-103
	1000	28-148	23-127
	3000	32-163	32-167
	5000	34-162	37-202
WOLF x_0= -1	5	17-35	14-29
	100	44-89	44-99
	1000	50-101	64-129
	3000	192-401	176-364
	5000	113-243	110-327
	5	30-83	30-85
ROSEN	100	30-83	30-85
	1000	30-83	30-85
	3000	30-83	30-85
	5000	31-80	31-88
	5	11-28	11-28
	100	12-30	12-30
BEAL x_0= (0,0)	1000	12-30	12-30
	3000	12-30	12-30
	5000	12-30	12-30
	5	5-14	5-14
	100	5-14	5-14
EDGER x_0= (1,0)	1000	5-14	5-14
	3000	6-16	6-16
	5000	6-16	6-16
CUBIC x_0= (-1,2,1)	5	13-38	16-47
	100	13-38	16-45
	1000	13-38	16-45
	3000	13-38	16-47
	5000	13-38	16-45
	5	8-21	8-21
	100	9-24	9-24
Shallow	1000	9-24	9-24
	3000	9-24	9-24
	5000	9-24	9-24
	5	24-67	24-67
	100	24-67	24-67

Table 1 (Cont.): A comparative performance analysis between the new algorithm and the standard LS-CG method.

Non-Diagonal	1000	24-67	24-67
	3000	24-67	24-67
	5000	24-67	24-67
	5	9-26	9-26
	100	9-26	9-26
Fred	1000	9-26	9-26
	3000	9-26	9-26
	5000	9-26	9-26
	5	6-18	6-18
	100	6-18	6-18
Recip3	1000	6-18	6-18
	3000	6-18	6-18
	5000	6-18	6-18
Total		1159-3270	1286-3950

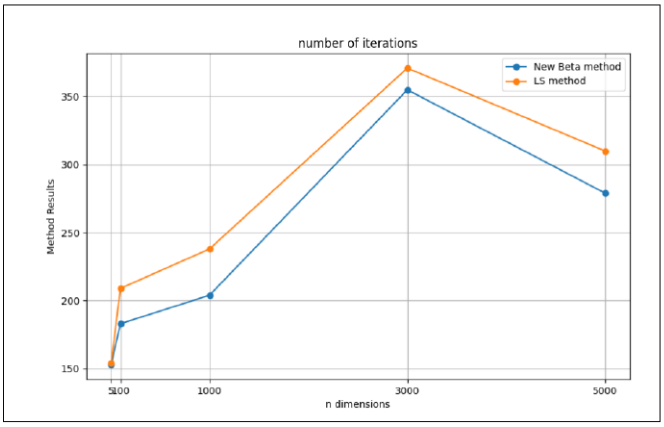

Figure 1: A comparative performance analysis between the new algorithm and the standard LS-CG method across the number of iterations (NOI).

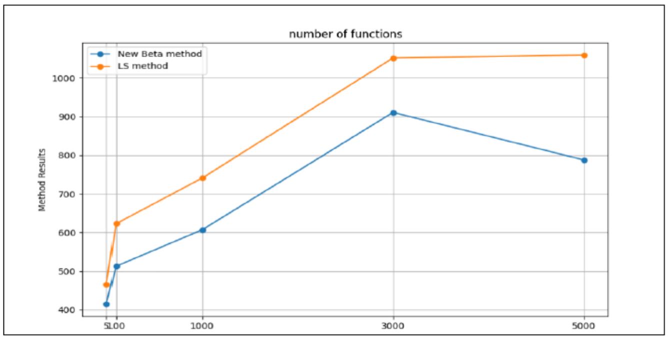

Figure 2: A comparative performance analysis between the new algorithm and the standard LS-CG method across the number of functions (NOF).

5. Conclusion

In the study, we introduced a new optimization technique and assessed its performance in comparison with LS-CG (Liu-Storey Conjugate Gradient). The algorithm that I proposed needs between 1159 and 3270 iterations and function evaluations, while the LS-CG method uses 1286 to 3950. These findings show that the novice computer programmer's method runs faster than the original method. The fact that both the number of iterations and function evaluations are lowered suggests that the improved CG technique improves speed and cuts running time. So, it is generally used for big optimization tasks.

On the whole, the approach appears to be a sufficient replacement for LS-CG because it reaches the same accuracy quickly and consumes less computational power. Ongoing efforts could develop adaptable ways to choose parameters and use this approach for dealing with the challenges discussed above.

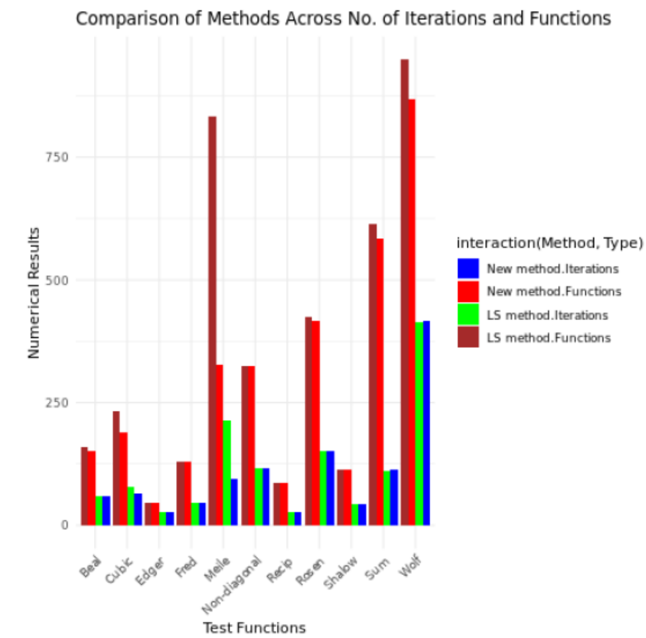

Figure 3: A comparative performance analysis between the new algorithm and the standard LS-CG method across the number of iterations (NOI) and functions (NOF).

References

Dai, Y. H., And Yuan, Y., 1999. A Nonlinear Conjugate Gradient Method With A Strong Global Convergence Property. Siam Journal On Optimization, 10(1), Pp. 177-182.

Daniel, J. W., 1967. The Conjugate Gradient Method For Linear And Nonlinear Operator Equations. Siam Journal On Numerical Analysis, 4(1), Pp. 10-26.

Fletcher, R., 2000. Practical Methods Of Optimization. John Wiley And Sons.

Fletcher, R., And Reeves, C. M., 1964. Function Minimization By Conjugate Gradients. The Computer Journal, 7(2), Pp. 149-154.

Hestenes, M. R., And Stiefel, E., 1952. Methods Of Conjugate Gradients For Solving Linear Systems (Vol. 49, No. 1). Washington, Dc: Nbs.

Ibrahim, A. L., Sadiq, M. A., And Shareef, S. G., 2019. A New Conjugate Gradient Coefficient For Unconstrained Optimization Based On Dai-Liao. Science Journal Of University Of Zakho, 7(1), Pp. 34-36.

Khatab, H. A., And Shareef, S. G., 2024. Two New Limited-Memory

Preconditioned Conjugate Gradient Algorithms For Nonlinear Optimization Problems. Journal Of Intelligent And Fuzzy Systems, (Preprint), Pp. 1-14.

Liu, Y., And Storey, C., 1991. Efficient Generalized Conjugate Gradient Algorithms, Part 1: Theory. Journal Of Optimization Theory And Applications, 69, Pp. 129-137.

Nocedal, J., And Wright, S. J. (Eds.)., 1999. Numerical Optimization. New York, Ny: Springer New York.

Polak, E., And Ribiere, G., 1969. Note Sur La Convergence De Méthodes De Directions Conjuguées. Revue Française D'informatique Et De Recherche Opérationnelle. Série Rouge, 3(16), Pp. 35-43.

Polyak, B. T., 1969. The Conjugate Gradient Method In Extremal Problems. Ussr Computational Mathematics And Mathematical Physics, 9(4), Pp. 94-112.

Shareef, S. G., 2022. A New Conjugate Gradient With Global Converges For Nonlinear Problems. Journal Of Duhok University, 25(2), Pp. 573-578.